Select Updates for Recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA) Waiver Applications for Manufacturers of In Vitro Diagnostic Devices

---

Draft Guidance for Industry and Food and Drug Administration Staff

DRAFT GUIDANCE

This draft guidance document is being distributed for comment purposes only.
Document issued on November 29, 2017.

You should submit comments and suggestions regarding this draft document within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to http://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. Identify all comments with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions about this document, please contact the Office of In Vitro Diagnostics and Radiological Health (OIR) at 301-796-5711, Marina Kondratovich, PhD, 301-796-6036 or by email at marina.kondratovich@fda.hhs.gov or Peter Tobin, PhD, 240-402-6169 or by email at peter.tobin@fda.hhs.gov.

When final, this document will update and supersede Section V. of the Recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA) Waiver Applications for Manufacturers of In Vitro Diagnostic Devices guidance, issued on January 30, 2008.

FDA U.S. FOOD & DRUG ADMINISTRATION
CENTER FOR DEVICES & RADIOLOGICAL HEALTH

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health
Office of In Vitro Diagnostics and Radiological Health

Preface

Additional Copies

Additional copies are available from the Internet. You may also send an e-mail request to CDRH-Guidance@fda.hhs.gov to receive a copy of the guidance. Please use the document number 16046 to identify the guidance you are requesting.

Table of Contents

Select Updates for Recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA) Waiver Applications for Manufacturers of In Vitro Diagnostic Devices

---

Draft Guidance for Industry and Food and Drug Administration Staff

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff or Office responsible for this guidance as listed on the title page.

I. Introduction and Scope

FDA has developed this draft guidance to implement section 3057 of the 21st Century Cures Act [P.L. 114-255], which requires FDA to revise "Section V. Demonstrating Insignificant Risk of an Erroneous Result — Accuracy" of the guidance "Recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA) Waiver Applications for Manufacturers of In Vitro Diagnostic Devices" ("2008 CLIA Waiver Guidance") that was issued on January 30, 2008. This draft guidance updates FDA's thinking regarding the appropriate use of comparable performance between a waived user and a moderately complex laboratory user to demonstrate accuracy. The 2008 CLIA Waiver Guidance remains in effect, in its current form, until this draft guidance is finalized, at which time the updates in Section III of this draft guidance will supersede the recommendations in Section V. of the 2008 CLIA Waiver Guidance.

FDA will incorporate the updates of the final version of this draft guidance into "Section V. Demonstrating Insignificant Risk of an Erroneous Result — Accuracy" of the 2008 CLIA

Waiver Guidance. The remainder of the 2008 CLIA Waiver Guidance will not be changed by this update and will remain in effect.

For the current edition of the FDA-recognized standard(s) referenced in this document, see the FDA Recognized Consensus Standards Database Web site at http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfStandards/search.cfm.

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. Background and Rationale

The Secretary of Health and Human Services has delegated to FDA the authority to determine whether particular tests are "simple" and have "an insignificant risk of an erroneous result" under CLIA and thus eligible for waiver categorization (69 FR 22849, April 27, 2004). The Centers for Medicare & and Medicaid Services (CMS) is responsible for oversight of clinical laboratories, which includes issuing Certificates of Waiver. CLIA requires that clinical laboratories obtain a certificate before accepting materials derived from the human body for laboratory tests. (42 U.S.C. § 263a(b)). Laboratories that perform only tests that are "simple" and that have an "insignificant risk of an erroneous result" may obtain a certificate of waiver. (42 U.S.C. § 263a(d)(2)).

CLIA, 42 U.S.C. § 263a(d)(3) Examinations and Procedures, as modified by the Food and Drug Administration Modernization Act of 1997 (FDAMA), reads as follows regarding tests that may be performed by laboratories with a Certificate of Waiver:

> The examinations and procedures [that may be performed by a laboratory with a Certificate of Waiver]… are laboratory examinations and procedures that have been approved by the Food and Drug Administration for home use or that, as determined by the Secretary, are simple laboratory examinations and procedures that have an insignificant risk of an erroneous result, including those that — (A) employ methodologies that are so simple and accurate as to render the likelihood of erroneous results by the user negligible, or (B) the Secretary has determined pose no unreasonable risk of harm to the patient if performed incorrectly.

The 2008 CLIA Waiver Guidance describes recommendations for device manufacturers about study design and analysis for CLIA Waiver by Application to support an FDA determination as to whether the device meets the statutory criteria for waiver described above. This update provides additional details and pathways for demonstrating that a test has an insignificant risk of erroneous result which is a key element for obtaining a CLIA Waiver by Application.

In developing this specific update, we have considered interactions with stakeholders since the issuance of the final guidance on January 30, 2008.

III. Revised Section V.

A. Demonstrating Insignificant Risk of an Erroneous Result – "Accuracy"

As stated previously, a CLIA waiver can be granted for, among others, tests that are "simple laboratory examinations and procedures that have an insignificant risk of an erroneous result" (42 U.S.C. § 263a(d)(3)). This includes tests that employ methodologies that are "so simple and accurate" that the "likelihood of an erroneous result by the user" is rendered "negligible" (42 U.S.C. § 263a(d)(3)(A)). One of the key elements for granting a CLIA waiver is that the test is accurate.

Accuracy is a widely used and generally understood term in clinical laboratory science. For example, the Clinical Laboratory Standard Institute in EP21[1] defines accuracy as "the closeness of an agreement between a test result and the accepted reference value." Accuracy takes into account the random and systematic components of error. The more accurate a test is, the less error there is, and therefore, the more likely the test is to have an "insignificant risk of erroneous result." With this in mind, there are various ways that a diagnostic test can be demonstrated to be accurate so that it can be granted a CLIA waiver by application.

For purposes of this draft guidance and the 2008 CLIA Waiver Guidance, the following terms are defined as:

- *Untrained Operator or Waived User*: An operator in waived settings with limited or no training or hands on experience in conducting laboratory testing.

- *Trained Operator or Moderate Complexity Laboratory User*: A laboratory professional who meets the qualifications to perform moderate complexity testing and with previous training in performing the test.

Sponsors developing devices designed for the CLIA-waived setting have traditionally taken a sequential route, first obtaining FDA clearance or approval and then submitting data for CLIA waiver determination. In 2012, the MDUFA III Commitment Letter created a single submission pathway referred to as a Dual 510(k) and CLIA Waiver application in which an applicant has an option to apply for marketing authorization and CLIA waiver concurrently within one submission. Proposed recommendations for Dual submissions are provided in the draft guidance "Recommendations for Dual 510(k) and CLIA Waiver by Application Studies"

[1] CLSI. *Evaluation of Total Analytical Error for Quantitative Medical Laboratory Measurement Procedures*. 2nd ed. CLSI document EP21; Wayne, PA: Clinical and Laboratory Standards Institute, 2016.

https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM586502.pdf, which, when finalized, would represent FDA's current thinking on recommendations for Dual submissions. This draft guidance outlines two options for a sequential route for a CLIA waiver in which safety and effectiveness or substantial equivalence is established first, followed by a separate application containing an evaluation of the test's accuracy in a CLIA-waived setting:

> **Option 1**: If the sponsor chooses to demonstrate the accuracy of the test when performed by trained operators as part of the marketing submission through comparison to a traceable calibration method (or reference method)[2], the sponsor can leverage these data in combination with a new study to demonstrate agreement between results of the test performed by untrained operators and trained operators in the waiver by application submission. Valid statistical methods may then be used to estimate the accuracy of the test when performed by untrained operators using a combined analysis of the data from both studies.

> **Option 2**: If the sponsor chooses to demonstrate substantial equivalence or safety and efficacy for the test when performed by trained operators in the marketing application without demonstrating accuracy through a comparison to a traceable calibration method (or reference method), the sponsor should demonstrate accuracy of the test when performed by untrained operators through direct comparison to a traceable calibration method (or reference method), or other comparative method performed in a laboratory setting by trained operators in the waiver application.

Choosing an option:

In many cases, a premarket submission for a test contains information about the test performed by laboratory professionals as compared to a reference or traceable calibration method, which allows FDA to assess the test's performance with trained operators. Under Option 1, a second comparison is then made to assess the level of agreement between the performance of trained and untrained operators. Because Option 1 includes two comparisons in a step-wise fashion, it introduces more potential sources of bias and higher levels of uncertainty (e.g., an increase of confidence intervals, potential bias due to differences in patient populations between studies) in the estimation of the candidate test's performance in the hands of the untrained operators as compared to Option 2. For quantitative tests, for example, data analysis of the agreement study between test results in the hands of untrained and trained operators in Option 1 provides only an estimation of a difference in systematic errors of the test in the hands of untrained and trained operators at medical decision levels. This systematic difference presents an additional source of error in the accuracy estimates of the test in the hands of untrained operators. However, if a quantitative test's performance in the hands of trained operators is sufficiently good, there may be sufficient allowable error left to accommodate the additional bias associated with this

[2] CLSI. *Metrological Traceability and Its Implementation; A Report.* CLSI document EP32-R, Wayne, PA: Clinical and Laboratory Standards Institute, 2006.

approach. For qualitative tests, agreement studies (i.e., comparison of the performance of the test by an untrained operator versus the performance of the test by a trained operator) in Option 1 typically demonstrate levels of agreement less than 100%, and applicants should consider whether the accuracy demonstrated in the marketing application and the expected agreement are high enough such that a combined statistical analysis will support that the test when performed by untrained operators is accurate. For those sponsors that want the best study to determine accuracy, the study design described in Option 2 is preferred for all types of tests (qualitative, semi-quantitative and quantitative), even when accuracy was determined in the marketing application with trained users.

In the case of certain tests, a direct estimation of the accuracy of the test in the hands of trained users may not be provided in the pre-market submission. For example, for moderate-risk tests, FDA premarket review evaluates whether the test when performed by laboratory professionals has substantially equivalent performance to that of a predicate device and may not include a comparison to a reference or traceable calibration method. If data are not available to support accuracy of the test with trained operators compared to a reference or traceable calibration method, these tests should use Option 2 for obtaining CLIA waiver by application. Option 1 is generally not appropriate for this scenario because, without some connection to a reference or traceable calibration method or initial understanding of a test's accuracy, it is not possible to determine the likelihood of an erroneous result of the test in the hands of untrained operators by only comparing results of the test in the hands of trained and untrained operators.

If using Option 1, applicants should provide the following summary information regarding the study, demonstrating the accuracy of the test when performed by trained operators, which they are leveraging:

- The number and description of sites in the comparison study, and the number and description of trained operators per individual site.

- The number of samples in the study, whether these samples were patient samples (if so, provide inclusion/exclusion criteria), and if some samples were archived or contrived, the number of such samples and explanation of how these samples were selected or prepared. The study subjects/samples should be representative of the intended use population at CLIA waived sites.

- The duration of the comparison study and the sources of variability that were included (e.g., different days, different runs, different lots and so on).

- The type of comparator method (CM) used in the study (e.g., traceable calibration method, reference method, gold standard, etc.). If the CM is a quantitative traceable calibration method or qualitative traceable cutoff method, provide information supporting traceability of calibrators or cutoffs, as appropriate.

- Statistical analysis, depending on the type of assay:

- For quantitative assays, provide the linear regression equation for the test results with trained operators (Y-axis) versus CM results with trained operators (X-axis) as $Test_{Trained} = A_1 \cdot CM + A_0$, the 95% confidence intervals for the slope A_1 and the intercept A_0 and the estimates of biases at the medical decision levels along with 95% confidence intervals.

- For qualitative assays, provide clinical/analytical performance of the test in the hands of trained operators as sensitivity ($Se_{Trained}$) and specificity ($Sp_{Trained}$) (or positive percent agreement ($PPA_{Trained}$) and negative percent agreement ($NPA_{Trained}$)), likelihood ratios, and other applicable measures of performance with 95% two-sided confidence intervals.

Manufacturers may submit a Pre-Submission to obtain feedback from FDA on planned study designs prior to conducting the study. This may be especially helpful if alternate approaches are planned. For additional information on the Pre-Submission process, please refer to FDA's guidance "Requests for Feedback on Medical Device Submissions: The Pre-Submission Program and Meetings with Food and Drug Administration Staff" available at https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm311176.pdf.

Regardless of the option selected, how accuracy is determined will be highly influenced by whether the test is quantitative, qualitative or semi-quantitative.

1) Quantitative Tests

A quantitative test is a test that gives numerical results (e.g., concentration of an analyte in a patient sample) which are referenced to a measuring interval and standards. A quantitative test is considered accurate for the purposes of CLIA waiver if the measurement results obtained by untrained operators in CLIA-waived testing settings are comparable to the results of an established quantitative test (i.e., for which results have been traced to references of higher order, usually national or international standards) performed by laboratory professionals. Determination of the comparability of the test results by untrained operators may be made directly (Option 2), or stepwise (Option 1), as long as a conclusion can be made that the candidate test has a negligible likelihood of erroneous results by untrained operators.

In order to define the accuracy of the candidate test one needs to define the relationship of the value obtained by the test to the true value, and define what contributes to any difference in the values. The following are possible CMs that can be used to determine the accuracy of the candidate test:

- Reference Method: The preferred CM is a reference method which has been thoroughly investigated and shown to yield accurate and precise results.
- Traceable Calibration Method: If a reference method is not available, an applicant might consider a traceable calibration method that includes calibrators traceable to references of higher order. A reference of higher order can be a certified reference material, a reference measurement procedure, or a network of reference

laboratories. The results of a traceable calibration method can be related to stated references, usually national or international standards, through an unbroken chain of comparisons in which measurement uncertainties have been documented at every step in the procedure.

Applicants who choose a traceable calibration method should provide information supporting traceability, including a description of reference materials or methods used to establish the traceability of the calibrators, the values and uncertainties for each of those reference materials or methods, and any other relevant information concerning traceability.

- Other Well-Documented Method: If neither a reference method nor a traceable calibration method are available, then an applicant may consider another well-documented method. In some situations, the test in the hands of trained operators can be considered as a CM (for example, if the candidate test in the hands of trained operators was evaluated versus a clinical gold standard for the presence/absence of the target condition and there were no other cleared/approved tests on the market).

In general, the inaccuracy of the test when performed by untrained operators is the deviation (error) of a single measurement from the true value which is obtained by traceable calibration method (or reference method). This includes both the systematic error described by bias and all sources of random errors such as imprecision and random interferences (i.e., patient sample-related random bias). Total error is the largest absolute possible error with 95% confidence and includes all random and systematic errors that can occur during the total testing process from the specimen collection through pre-analytical, analytical and post-analytical phases, see Figure 1.

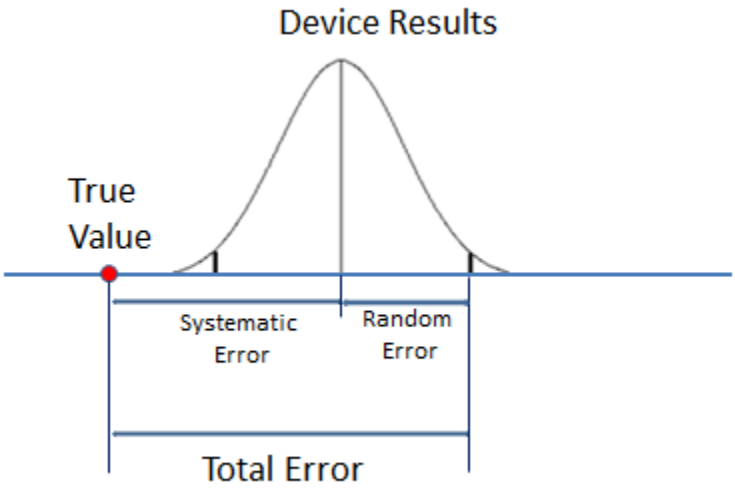

Figure 1. Systematic error, random error and total error.

All tests have bias and random error, so one needs to define whether the deviation of a single measurement from true value (error) that includes bias and random error is clinically significant.

A way to incorporate that into the statistical analysis is by defining zones of allowable total error where 95% of the test values should fall, and zones where no test values should fall, zones of limits of erroneous results. Defining the limits on the allowed error incorporates into the numerical analyses clinical judgment on the benefits and risks posed by the test when used as intended.

- *The Allowable Total Error (ATE) zones.*
  The ATE zones represent a range of values that can be tolerated without invalidating the medical usefulness of the test. Typically at least 95% of the sample results should fall within the established ATE zones.

- *The Zones of Limits for Erroneous Results (LER).*
  While it may be acceptable for a small percentage of results to fall outside the ATE zone, such results should remain within wider limits. Accordingly, applicants should specify Limits of Erroneous Results (LER) zones, within which it is not acceptable for any results to occur as this would pose a risk to patient safety.

When graphed, such zones give you an error grid as seen in Figure 2.

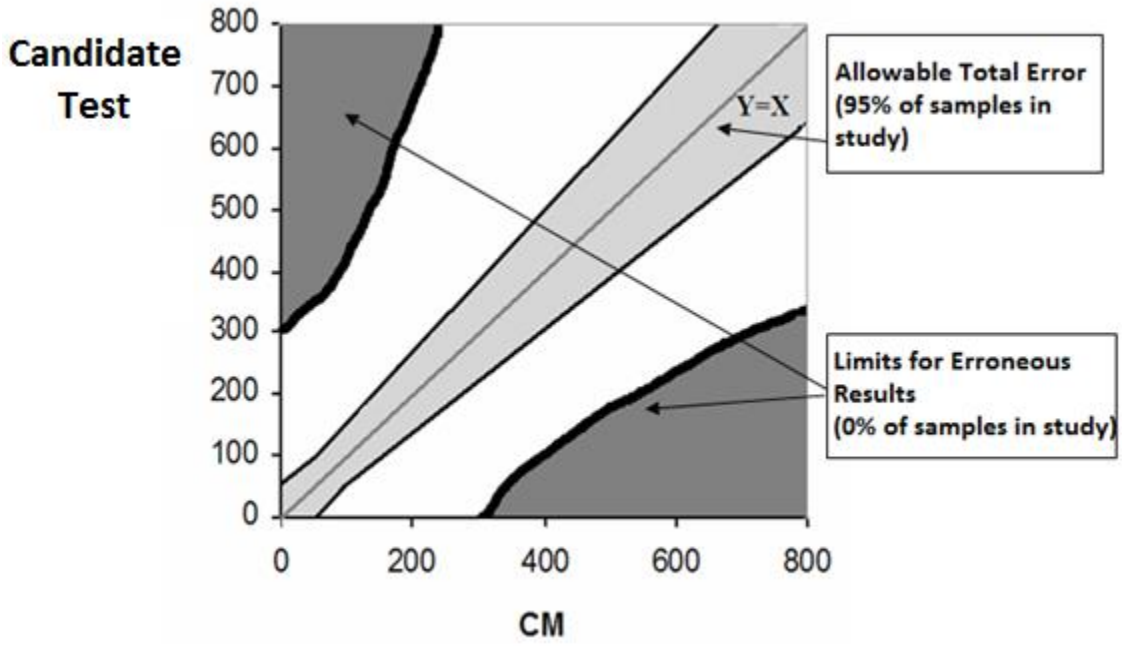

Figure 2. Example of ATE and LER zones.

Applicants should establish zones of ATE and LER prior to initiating a study.

For Option 1, in the agreement study, each of the subjects should be tested by the candidate test twice: once by untrained operators and once by trained operators. These data are used for estimation of systematic difference between test results by untrained operators versus trained

operators at medically important levels (see Figure 3 below). In addition, evaluation of imprecisions of the test when run by untrained and trained operators should be performed for three concentrations (low, medium, and high) and these data are used for comparing test imprecisions in the hands of untrained and trained operators.

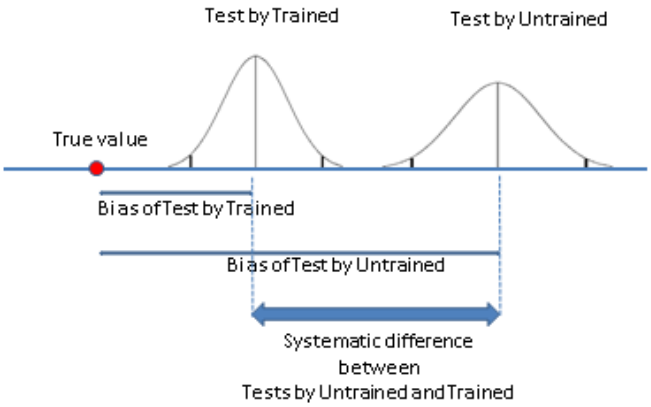

Figure 3. Systematic difference between untrained and trained operators

For Option 2, possible study designs and analysis plans for quantitative tests should align with the Clinical and Laboratory Standards Institute (CLSI) documents EP21 "Evaluation of Total Analytical Error for Quantitative Medical Laboratory Measurement Procedures."[3]EP09[4] and EP17[5]. The measuring interval of the CM should be at least as wide as the measuring interval of the candidate test and the CM should be run in laboratory settings by trained operators.

For both options, testing samples across the measuring interval of the candidate test in the study is recommended.

FDA recommends that applicants interact with FDA on their proposed study design, establishing ATE and LER through the Pre-Submission process prior to initiating a study.

2) Qualitative Tests

A qualitative test is a test that provides two outputs (e.g., positive/negative or yes/no) or multiple nominal[6] categories. A qualitative test is considered accurate for the purposes of CLIA waiver if the results obtained by untrained operators in CLIA waived settings are comparable to the results of a gold standard (i.e., the best available method for determining whether a target condition is present or absent), or to the results of a quantitative traceable calibration test with an appropriate

[3] CLSI. *Evaluation of Total Analytical Error for Quantitative Medical Laboratory Measurement Procedures*. 2nd ed. CLSI guideline EP21. Wayne, PA: Clinical and Laboratory Standards Institute, 2016.
[4] CLSI. *Measurement Procedure Comparison and Bias Estimation Using Patient Samples; Approved Guideline* – Third Edition. CLSI document EP09-A3, Wayne, PA: Clinical and Laboratory Standards Institute, 2013.
[5] CLSI. *Evaluation of Detection capability for Clinical Laboratory Measurement procedures; Approved Guideline* – Second Edition. CLSI document EP17-A2, Wayne, PA: Clinical and Laboratory Standards Institute, 2012
[6] Nominal categories are categories with no intrinsic ordering. For example, an IVD test for genotyping HCV that gives results of multiple categories as 1a, 1b, 2, 3, 4, 5 and 6 is a qualitative test.

cutoff, or to the results of a qualitative traceable cutoff test (i.e., when cutoff values have been traced to references of higher order) performed in a laboratory setting by trained operators. Determination of the comparability of the test results by untrained operators may be made directly (Option 2), or stepwise (Option 1), such that a conclusion can be made that the candidate test has a negligible likelihood of erroneous results by untrained operators.

For qualitative tests, information about bias and random error may not be directly obtainable. In general, the accuracy of a qualitative test is determined by comparison of the test results with one of the comparative methods (CM) described below:

- **Gold Standard or Qualitative Reference Method**: The preferred CM is a gold standard (a best available method for determining whether the target condition is present or absent (for example, composite reference method, a clinical diagnosis determined by definitive clinical methods), or a qualitative reference method for determination of the presence or absence of the analyte of interest (for examples of qualitative reference methods, consult with OIR)).

- **Quantitative Reference Method**: a quantitative reference method such as those outlined in the section on quantitative tests, with the appropriate cutoff value for positive and negative results.

- **Traceable Calibration Method**: A quantitative traceable calibration method with measurement results traceable to higher-order references such as that outlined in the section on quantitative tests with the appropriate cutoff value for positive and negative results.

- **Traceable Cutoff Method**: A qualitative traceable cutoff method with cutoff determination traceable to higher-order references.

- **Other Well-Documented Method**: If neither a reference method nor a traceable calibration method are available, then an applicant may consider another well-documented quantitative or qualitative method. For example, this could be a method that was tested by reference specimen panels (e.g., panels of samples prepared by well recognized institutions, such as WHO, CDC, NIST).

If a gold standard is not available, the performance of a qualitative test may be assessed relative to a quantitative CM with an appropriate cutoff for determination of positive and negative results by the quantitative CM. The quantitative CM should either be a reference method or a traceable calibration method. If such a CM is not available, the applicant may consider a qualitative traceable cutoff method or other well-documented method, such as use of reference specimen panels. We recommend discussing alternatives to quantitative CMs with FDA through the Pre-Submission process prior to initiating such a study.

In general, accuracy of a qualitative test with two outputs is described by the performance metrics presented below based on the test results and gold standard results for N representative subjects from the target population:

Table 1. Qualitative Test vs. Gold Standard

		Gold Standard		Total
		Target Condition present	Target Condition absent	
Test	Pos	A	B	A+B
	Neg	C	D	C+D
Total		N_1=A+C	N_0=B+D	N

- Clinical Sensitivity (Se) = A/(A+C)
- Clinical Specificity (Sp) = D/(B+D)
- Positive Likelihood Ratio (PLR) = $(A/N_1)/(B/N_0)$
- Negative Likelihood Ratio (NLR) = $(C/N_1)/(D/N_0)$
- Prevalence (π)[7] = N_1/N
- Positive Predictive Value (PPV) = A/(A+B)
- Negative Predictive Value (NPV) = D/(C+D)
- Percent of subjects with Test positive results = (A+B)/N
- Percent of subjects with Test negative results = (C+D)/N

If the gold standard is not available, test performance metrics are positive and negative percent agreements described below based on the test and comparative method results for N representative subjects from the target population:

Table 2. Qualitative Test vs. Quantitative Test

		CM		Total
		Positive	Negative	
Test	Pos	A	B	A+B
	Neg	C	D	C+D
Total		N_1=A+C	N_0=B+D	N

- Positive Percent Agreement (PPA) = A/(A+C)
- Negative Percent Agreement (NPA) = D/(B+D)

The formulas for sensitivity, specificity, and positive and negative percent agreements appear to be the same; however, it is critical to recognize that these values are not interchangeable[8].

[7] If the prevalence in the study was different from the prevalence in the target population, then PPV and NPV for this prevalence can be calculated using the following relationship between predictive values, likelihood ratios and prevalence as: $\frac{PPV}{1-PPV} = PLR \cdot \frac{\pi}{1-\pi}$ $\frac{PPV}{1-PPV} = PLR \cdot \frac{\pi}{1-\pi}$ and $\frac{1-NPV}{NPV} = NLR \cdot \frac{\pi}{1-\pi}$ $\frac{1-NPV}{NPV} = NLR \cdot \frac{\pi}{1-\pi}$.

[8] CLSI. *User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline – Second Edition*. CLSI document EP12-A2, Wayne, PA: Clinical and Laboratory Standards Institute, 2008.

Applicants should establish criteria for acceptable levels of performance prior to beginning a study regardless of the CM selected. Criteria should be based on a benefit-risk analysis taking into consideration benefits for patients (e.g., availability of results at the time of the doctor's office visit) and risks (e.g., incorrect results when performed by untrained operators). FDA generally recommends that criteria for sensitivity (or PPA between test and CM) and specificity (or NPA between test and CM) should be 95% or greater; and that the lower confidence bound for the two-sided 95% score confidence interval should be at least 89%. In some cases, tighter performance criteria may be needed to reasonably assure that the candidate test is "accurate" and in other cases lower performance criteria may be clinically acceptable with sufficient benefit-risk justification.

For Option 1, in the agreement study, each of the subjects should be tested by the candidate test twice: once by untrained operators and once by trained operators. These data are used for calculation of PPA and NPA between test performed by untrained operators and test performed by trained operators:

Table 3. Test by Untrained Operators vs. Test by Trained Operators

		Test by Trained		Total
		Positive	Negative	
Test by Untrained	Pos	A	B	A+B
	Neg	C	D	C+D
Total		N_1=A+C	N_0=C+D	N

Estimate the $PPA_{Untrained|Trained}$ and the $NPA_{Untrained|Trained}$ as

$PPA_{Untrained|Trained}$=A/(A+C)

$NPA_{Untrained|Trained}$=D/(B+D)

If a numeric signal of the test is available, then you can calculate zones around the cutoff for which samples may have discordant results if run multiple times by trained operators. Using these zones, we recommend that you present results of the agreement study in a detailed table of agreement:

Table 4. Test by Untrained Operators vs. Numeric Signal of Test by Trained Operators

		Test by Trained				
		Positive		Negative		Total
		High and Moderate Positive	Low Positive (close to cutoff)	High Negative (close to cutoff)	Low and Moderate Negative	
Test by Untrained	Pos					
	Neg					
Total		N_1		N_0		N

Taking into consideration acceptable levels of sensitivity (or PPA between test and CM) and specificity (or NPA between test and CM) and levels of test accuracy in the pre-market submission, you should evaluate whether the levels of agreement between test results in the hands of untrained and trained operators in the agreement study are acceptable.

Usually, an agreement study does not have enough subjects close to the cutoff that test performance with untrained and trained operators for the subjects close to the cutoff can be evaluated. Therefore, in addition, comparison of imprecisions of the test when run by untrained and trained operators should be performed for concentrations C_5 and C_{95} [9] (or limit of detection) close to the cutoff.

For Option 2, the study should include subjects from the intended use population and these subjects should be tested by untrained operators and results should be compared with results from the CM performed by trained operators. In addition, samples near the cutoff (i.e., C_5 and C_{95} of the CM if the CM is a quantitative test or C_5 and C_{95} of the candidate test if the CM is a gold standard for the target condition) should be tested by untrained operators. For analysis plans for qualitative tests, see the Clinical and laboratory Standards Institute (CLSI) document EP12.[10]

For Options 1 and 2, we recommend discussing study sample with FDA through the Pre-Submission process prior to initiating a study.

3) Semi-Quantitative Tests

For the purpose of this guidance, a semi-quantitative test is a test with a few ordinal categories (e.g., negative, trace, +, ++, +++) in which the order of categories together with the definitions of these categories contain information used during the interpretation of the test results. A semi-quantitative test is considered accurate for the purposes of CLIA waiver if the results obtained by untrained operators in CLIA-waived testing settings are comparable to the results obtained by trained operators performing a traceable calibration test (see "Quantitative Tests" section above for more information), with appropriate cutoffs for defining ordinal categories

An example of a semi-quantitative test with four ordinal categories (negative, trace, +, ++) is presented in Figure 4.

[9] CLSI. *User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline – Second Edition*. CLSI document EP12-A2, Wayne, PA: Clinical and Laboratory Standards Institute, 2008.
[10] CLSI. *User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline – Second Edition*. CLSI document EP12-A2, Wayne, PA: Clinical and Laboratory Standards Institute, 2008.

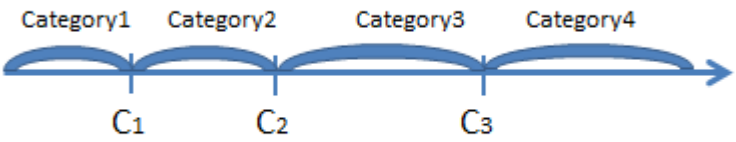

Figure 4. Semi-quantitative test with 4 categories based on three (3) cutoffs for numeric values.

The performance of a semi-quantitative test may be assessed relative to a quantitative comparative method (CM). The quantitative CM should either be a reference method or a traceable calibration method. Any quantitative CM chosen will have some random error associated with it such that two measurements of the same sample may differ slightly. Thus, a sample close to a cutoff C of the candidate semi-quantitative test may have an initial result above the cutoff C and subsequent result below the cutoff C. As samples close to cutoff C may have discordant results when tested multiple times by the quantitative CM, one can anticipate that these samples may also have discordant categories when tested by the candidate semi-quantitative test. A zone[11] around each cutoff can be established based on the allowable imprecision of the quantitative CM (%CV or SD) such that 95% of the differences between two repeated measurements of the quantitative CM, X_2 and X_1, are inside this zone (see Figure 5).

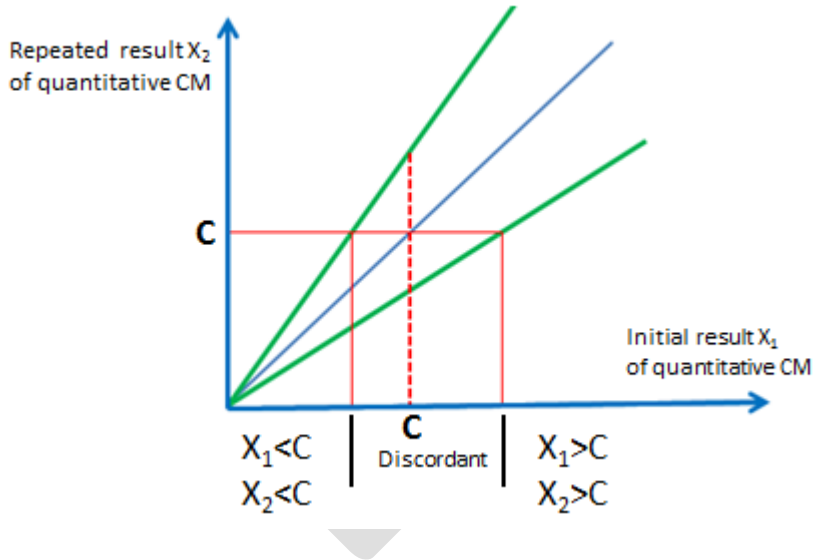

Figure 5. Two repeated measurements from quantitative test.

Applicants should establish near-cutoff ranges around each cutoff using the 95% zone as described above. For samples with measurements by the CM within one of these ranges, it is expected that candidate semi-quantitative test results may fall into either of the categories bordering the cutoff. For CM test results outside of the near-cutoff ranges only one

[11] These boundaries are symmetrical around the diagonal when the distance is calculated perpendicular to the diagonal.

corresponding bin of the semi-quantitative test is allowed. To summarize, the ATE zones for semi-quantitative candidate tests are the allowable bins for the different ranges of the quantitative CM values. For CM results in the near-cutoff range, both semi-quantitative candidate test bins are allowable and only one corresponding bin of the semi-quantitative candidate test is allowable for the CM values outside of these near-cutoff ranges. The LER zones are defined as the bins of the semi-quantitative candidate test that differ from the allowable bins by more than one bin.

This analysis with ATE and LER zones may be presented as a table of agreement between semi-quantitative and quantitative tests as shown in Table 5 where the light grey and dark grey shaded cells represent the ATE and LER zones correspondingly.

Table 5. Agreement of Semi-Quantitative Test and Quantitative CM

		Quantitative CM							
			Near cutoff C_1		Near cutoff C_2		Near cutoff C_3		Total
Expected Test Categories		Only Cat.1	Cat.1 or Cat.2	Only Cat.2	Cat.2 or Cat.3	Only Cat.3	Cat.3 or Cat.4	Only Cat.4	
Test	Cat.1	A_1	B_1	C_1	D_1	E_1	F_1	G_1	
	Cat.2	A_2	B_2	C_2	D_2	E_2	F_2	G_2	
	Cat.3	A_3	B_3	C_3	D_3	E_3	F_3	G_3	
	Cat.4	A_4	B_4	C_4	D_4	E_4	F_4	G_4	
Total		A	B	C	D	E	F	G	N
Observed percent in ATE		A_1/A	$(B_1+B_2)/B$	C_2/C	$(D_2+D_3)/D$	E_3/E	$(F_3+F_4)/F$	G_4/G	
		Total percent in ATE= $(A_1+B_1+B_2+C_2+D_2+D_3+E_3+F_3+F_4+G_4)/N$ Total percent in LER= $(A_3+A_4+B_4+C_4+E_1+F_1+G_1+G_2)/N$							

For Option 1, in the agreement study, each of the subjects should be tested by the candidate test twice: once by untrained operators and once by trained operators. These data are used for calculation of percent agreements between test performed by untrained operators and test performed by trained operators. For example, for N samples from the intended use population in the agreement study, the performance of a semi-quantitative candidate test with four categories may be presented as described in Table 6. The grey shaded cells in Table 6 represent samples for which there was agreement between the test results by untrained and trained operators.

Table 6. Comparison of Two Semi-Quantitative Tests

		Test by Trained				
		Category1	Category2	Category3	Category4	Total

Test by Untrained	Category1	A_1	B_1	C_1	D_1	
	Category2	A_2	B_2	C_2	D_2	
	Category3	A_3	B_3	C_3	D_3	
	Category4	A_4	B_4	C_4	D_4	
Total		$A_1+A_2+A_3+A_4$	$B_1+B_2+B_3+B_4$	$C_1+C_2+C_3+C_4$	$D_1+D_2+D_3+D_4$	N

Performance of the semi-quantitative test is described by the percent agreements (PA) of the categories of the test in the hands of untrained operators with categories of the test in the hands of trained operators as:

- $PA_{Category1,Untrained|Categoryl,Trained} = A_1/(A_1+A_2+A_3+A_4)$
- $PA_{Category2,Untrained|Category1,Trained} = A_2/(A_1+A_2+A_3+A_4)$
- $PA_{Category3,Untrained|Category1,Trained} = A_3/(A_1+A_2+A_3+A_4)$
- $PA_{Category1,Untrained|Category2,Trained} = B_1/(B_1+B_2+B_3+B_4)$ and so on.

For Option 2, the study includes subjects from the intended use population and these subjects should be tested by untrained operators and results should be compared with results from the CM performed by trained operators. The CM should be either a reference method or a traceable calibration method. ATE and LER should be established and analysis should be performed as described above. We recommend discussing selection of the CM with FDA through the Pre-Submission process prior to initiating such a study.

For Options 1 and 2, in addition, samples near each of the clinically important cutoffs should be tested. The number of samples needed will depend on the number of categories and the importance of each category. For Option 1, the imprecisions of the candidate test when performed by untrained and trained operators should be compared by testing C_5 and C_{95} samples for each clinically important cutoff. For Option 2, samples close to the cutoffs (i.e., near C_5 and C_{95} of quantitative CM) should be tested by untrained operators for clinically important cutoffs.[12]

It should be noted that when a semi-quantitative test is compared to the same or another semi-quantitative test in a study and the level of agreement is not high, this study design does not provide enough information to evaluate whether the levels of disagreement are acceptable. This is because one cannot determine whether the samples with discordant results were close to the cutoffs (i.e., where some level of disagreement would be expected) or far from the cutoffs (i.e., where a very high level of agreement is anticipated). Accordingly, Option 1is recommended only when a quantitative CM is not available and a high level of agreement of the test results between untrained and trained operators is expected.

For Options 1 and 2, we recommend discussing study size with FDA through the Pre-Submission process prior to initiating a study.

[12] CLSI. *User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline – Second Edition*. CLSI document EP12-A2, Wayne, PA: Clinical and Laboratory Standards Institute, 2008.

4) General Study Design Considerations

For all study designs, FDA recommends that applicants evaluate test performance in settings designed to replicate, as closely as possible, the actual CLIA waived settings and patients/samples. Therefore, study designs should include the following:

- CLIA Waiver testing sites that are representative of the intended use population.
- CLIA Waiver intended operators (untrained operators). We encourage you to enroll operators with the least amount of training that might be encountered at the types of sites for which this device is intended.
- Intended sample type and matrix.
- Testing over time, as in the typical intended use setting.

a) Testing sites

You should conduct the study to support CLIA waiver at a minimum of three sites that are representative of both the intended use patient population and the intended operators in CLIA waived settings. Generally, the sites should include different demographic locations (e.g., outpatient clinic, physician's office), since patient populations and intended operators typically vary between different demographic locations. In your CLIA waiver application, you should present a brief description of each site, including its name, address, and the date the study was performed. If there were sites that were included at the beginning, but then did not complete the study, you should provide a brief explanation for why those sites did not complete the study.

b) Study participants

1. Operators

a) CLIA Waiver intended operators (untrained operators)

The study should include 1-3 untrained operators at each site and at least nine (9) untrained operators across all sites. You should ensure that the untrained operator study participants enrolled represent anticipated operators of the device you propose for CLIA waiver. We recommend that you record and tabulate the education (including experience and training) and the occupation of each operator to demonstrate that these participants meet the definition of intended operators and include this in your CLIA waiver application. In addition, for each study site, we ask you to report the same information on other operators that were available at the testing site but that were not chosen to participate.

b) Trained operators

The study should include at least one laboratory professional (trained operator) at each site. These laboratory professionals should meet the qualifications to perform moderate complexity testing and have previous training in performing the candidate test.

c) Instructions for use

You should provide the untrained operators who participate in the study with only the Quick Reference Instructions (see Section VI.C. Educational Information of the guidance entitled "Recommendations: Clinical Laboratory Improvement Amendments of 1988 (CLIA) Waiver Applications for Manufacturers of In Vitro Diagnostic Devices" (https://www.fda.gov/RegulatoryInformation/Guidances/ucm079632.htm). The untrained operators should receive no additional instructions (e.g., written or verbal training, coaching, or prompting). Likewise, untrained operators should have no opportunity to discuss the test with other participants or otherwise coach or observe each other. Untrained operators may call an 800 number help-line if such a service is to be provided for the device when it is marketed. You should include, in your waiver application, the instructions you provided to untrained operators participating in the study.

d) Universal precautions

You should comply with the Federal Food, Drug, and Cosmetic Act (FD&C Act) and its implementing regulations and should ensure your study complies with all other pertinent laws and regulations, including Occupational Health and Safety Administration (OSHA) regulations pertaining to biological hazards ("universal precautions"), 29 CFR 1910.1030.

e) Operator questionnaire

You should develop an operator questionnaire to be filled out by all untrained operators participating in the study. This questionnaire should be designed to help assess whether the untrained operators understood how to use the device correctly. It is important that the questionnaire be given to test untrained operators *after* the completion of the clinical study, so the questions do not bias the untrained operators during the study. Some questions may ask untrained operators to indicate agreement on a 1-5 scale (1=strongly disagree; 5=strongly agree). The following are examples:

- The instructions were easy to follow.
- It was easy to apply the sample correctly.
- It was easy to see and understand the test results (e.g., appearance of the line, change of color).
- The control line was always distinct and easy to read.

- The instructions clearly explain what to do if a test result does not appear or is invalid.
- I needed help from someone the first time I ran the test.

We recommend that, as part of the questionnaire, you show various possible test results and control results that are positive, negative, and invalid and ask the untrained operator to read these results. You may wish to present these questions as true/false or multiple choice questions.

You should also strongly encourage general comments by the untrained operators. We recommend that you include your survey questions and results with your CLIA application.

c) Subjects (Patients) and Samples

In order to prevent biases, samples should be collected from consecutive patients over one month. Depending on the specific clinical site, the prevalence of the disease, or other factors, it may be appropriate to limit consecutive enrollment to two (2) weeks.

You should ensure that subjects from whom you will obtain specimens for the clinical study meet inclusion and exclusion criteria corresponding to the intended use population of the test. Once a subject has been determined to meet appropriate inclusion criteria, he/she should be informed of the study and invited to participate.

You should follow applicable laws and regulations for human subject protection, including patient privacy and informed consent. See section 520(g) of the FD&C Act; 21 CFR parts 50,56, and 812; and the Health Insurance Portability and Accountability Act (HIPAA) [P.L. 104-191]; 45 CFR Part 46.

Samples should adequately represent all possible values of the test. If possible, applicants should strive to achieve this at each site as well as across all sites. For quantitative and semi-quantitative candidate tests, samples should span the measuring intervals of the device and study data should include a few samples around Medical Decision Levels (MDLs). For qualitative test, samples in the study should include samples near the cutoffs.

Each sample should be split in two parts. One part should be tested by the untrained operator using the test and the other part should be tested by a trained operator using the test for Option 1 and by CM for Option 2. If the sample cannot be split into parts, then a second sample from the same patient should be collected within a suitable time interval. We recommend consultation with FDA through the Pre-Submission process if the order in which the samples are collected impacts the results of testing. Untrained and trained operators should be blinded to test results from other operators and other samples.

We recommend using actual patient specimens to best assess a device in the hands of untrained operators. However, in some situations, when samples from some categories are rare, it may be appropriate to supplement prospective patient samples with archived samples. If archived patient samples are not available, it may be appropriate to supplement patient samples with

contrived samples, such as individual spiked or diluted patient samples. Spiked, diluted, or otherwise contrived samples used in the study should be individual samples (i.e., they should not be aliquots from a single pool). Any archived or contrived sample matrix should be the same as that of the intended use patient samples. Applicants should describe the origin of such samples and how they were prepared. For qualitative and semi-quantitative tests, archived and contrived samples should include samples near the cutoffs. Use of archived or contrived samples should be appropriately justified. In general, archived or contrived samples should not comprise greater than one third of the total study samples; however, there may be some situations in which more or less would be appropriate when an adequate justification is provided. The patient and contrived samples should be as equally distributed among the untrained operators as possible. FDA encourages applicants to discuss planned use of archived or contrived samples with OIR in advance.

d) Financial disclosure

If clinical investigators are involved in the clinical study, you should include a Financial Disclosure Statement with your waiver application. For information on financial disclosure statements, we recommend you consult, "Guidance for Industry: Financial Disclosure by Clinical Investigators," available at https://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM341008.pdf and 21 CFR Part 54, Financial Disclosure by Clinical Investigators.

e) Clinical study reports

You should report results of the clinical study intended to support CLIA waiver by each intended site and overall, if appropriate. Reports should include the following:

- Protocol description.
- Number of subjects (i.e., patients) studied.
- Procedures for subject inclusion and exclusion.
- Description of the subject population.
- Description of how specimens were collected and stored.
- Masking techniques.
- Discontinuations.
- Complaints, device failures, and replacements.
- Any invalid results and how these were handled.
- Information about QC procedures that were performed.
- Pertinent tabulations.
- Annotated line listings of results (including electronic versions).
- Clear descriptions and presentations of the statistical analyses.
- An explanation for data that are incomplete or missing (Note: You should not remove "outliers").

You should also report the following for each untrained and trained operator:

- Total number of performed candidate tests.
- Number of initial invalid results.
- Number of retested results.
- Number of final invalid results.

You should calculate and report the percentage of initial and final (if applicable) invalid results with a 95% two-sided confidence interval and then exclude invalid results from calculations of the test performance characteristics. Please provide a rationale as to why the observed percentage of invalid results is clinically acceptable.